

## Composition and Gene Expression of the *cag* Pathogenicity Island in *Helicobacter pylori* Strains Isolated from Gastric Carcinoma and Gastritis Patients in Costa Rica

ALESSANDRA OCCHIALINI,<sup>1</sup> ARMELLE MARAIS,<sup>1</sup> MARIA URDACI,<sup>2</sup> RAFAELA SIERRA,<sup>3</sup>  
NUBIA MUÑOZ,<sup>4</sup> ANTONELLO COVACCI,<sup>5</sup> AND FRANCIS MÉGRAUD<sup>1\*</sup>

*Laboratoire de Bactériologie, Université Victor Ségalen Bordeaux 2 et Hôpital Pellegrin, 33076 Bordeaux,<sup>1</sup>*

*Laboratoire de Microbiologie, Ecole Nationale d'Ingénieurs des Travaux Agricoles, 33170 Gradignan,<sup>2</sup>*

*and Centre International de Recherche sur le Cancer, 69372 Lyon,<sup>4</sup> France; University of  
Costa Rica, San José, Costa Rica,<sup>3</sup> and IRIS, 53100 Siena, Italy<sup>5</sup>*

Received 7 August 2000/Returned for modification 25 September 2000/Accepted 20 November 2000

**The composition and in vitro expression of the *cag* pathogenicity island genes in a group of *Helicobacter pylori* strains obtained from patients suffering from chronic gastritis-associated dyspepsia ( $n = 26$ ) or gastric carcinoma ( $n = 17$ ) were analyzed. No significant difference in the distribution of the 10 studied regions was found between the cases and the controls. Nine strains did not harbor any of the selected regions: eight (30.8%) isolated from patients with gastritis only and one (5.9%) from a patient with gastric carcinoma. No association was found between the number of repeated sequences at the 3' end of the *cagA* gene or the presence of tyrosine phosphorylation motifs and the clinical origin of the strains. The *virB10* homolog gene was the sole gene studied to be significantly expressed more often in cancer strains than in gastritis strains ( $P = 0.03$ ).**

Since the first culture of *Helicobacter pylori* in 1982, chronic infection with this bacterium has been identified as the etiological agent of gastritis and duodenal ulcer (18, 21). Evidence for a causal relationship between chronic *H. pylori* infection and stomach cancer came first from epidemiological studies (10, 19) and, more recently, from animal models of carcinogenesis (12, 32). In 1994, the International Agency for Research on Cancer-World Health Organization (11) defined *H. pylori* infection as a group I carcinogen (definite carcinogen).

Despite the very high rate of *H. pylori* infection in some populations, the rate of gastric cancer is relatively low. Beside environmental factors and hereditary predisposition, as recently illustrated with interleukin-1 $\beta$  polymorphism (8), a possible explanation may be found in the different patterns of pathogenicity of *H. pylori* strains. The *cag* pathogenicity island (PAI) is one of the major *H. pylori* virulence factors found more frequently in patients suffering from severe gastroduodenal diseases, including peptic ulcers (5, 13, 16) and gastric adenocarcinomas (16, 26).

Recent studies have shown that the CagA protein, encoded by the *cag* PAI, is translocated into host cell by the *cag* PAI type IV secretion system, by which it becomes phosphorylated by an unknown cell kinase (4, 25, 27, 31), inducing an active reorganization of actin (31). The sites of possible phosphorylation of CagA have been identified (25). The CagA protein varies in size from 128 to ca. 140 kDa, and this variation is related to the presence of a variable number of repeated sequences in the 3' region of the gene (9, 35, 36).

The aim of this study was to compare the presence, the composition, and the expression of several *cag* PAI genes,

including *cagA*, in Costa Rican strains obtained from patients suffering from gastric carcinoma ( $n = 17$ ) and from chronic gastritis-associated dyspepsia ( $n = 26$ ), as well as to determine the terminal sequence of the *cagA* gene and the putative tyrosine phosphorylation motifs (TPMs) of the deduced CagA protein.

**Molecular composition of the *cag* PAI among *H. pylori* strains.** After culture of the 43 *H. pylori* isolates on selective medium (20), the DNA was extracted using the standard phenol-chloroform procedure (24). Four reference strains were included in the study: J99 (3), 26695 (33), NCTC 11638 (2, 5), and SS1 (15). The presence of *cag* PAI was analyzed by detecting 10 different regions (the 5' and 3' ends of *cagA*, *cagC*, *virB4/cagF*, *cagL/cagN/cagM*, *cagP/cagQ/cagR*, *cagS/virB7*, *virB9*, *virB10*, *virB11*, and *virD4*) by PCR amplification and dot blot hybridization as previously described (24). The presence of IS605 (*tnpA* and *tnpB*) was also evaluated (Table 1). An example of *cagL/cagM/cagN* and IS605 detection by PCR is presented in Fig. 1A. As expected, a 1,803-bp fragment was revealed for the strains NCTC 11638, J99, and 26695, and no amplification was obtained in the control when DNA was replaced by water (lane NC). Among the 22 clinical strains presented in this figure, 15 yielded the expected fragment. In the case of IS605 amplification, a 1,486-bp amplicon was obtained in positive controls (26695 and NCTC 11638 strains), as well as in six strains among the 22 presented. As expected from the genome sequence, the IS605 fragment was not amplified from DNA of strain J99.

Negative results by PCR may be due either to an absence of the tested gene (true negative) or to a lack of primer annealing due to interstrain variation in the sequences targeted by the primers. For this reason, dot blot analyses were performed to confirm the absence of the selected genes by hybridization conducted under stringent conditions and as described elsewhere (24). Moreover, the specificity of the amplicons was also

\* Corresponding author. Mailing address: Laboratoire de Bactériologie, Hôpital Pellegrin, Place Amélie Raba-Léon, 33076 Bordeaux Cedex, France. Phone: 33-5-56-79-59-10. Fax: 33-5-56-79-60-18. E-mail: francis.megraud@chubordeaux.fr.

TABLE 1. Primer sequences and DNA amplification conditions for each selected region

Amplified region <sup>a</sup>	Primer	Primer sequence (5'→3')	Size of amplified fragment (bp)	Annealing step (temp [°C], time [min]) <sup>b</sup>
<i>cagA</i> (5' end)*†	D008	ATA ATG CTA AAT TAG ACA ACT TGA GCG A	298	51, 1.5
	R008	TTA GAA TAA TCA ACA AAC ATC ACG CCA T		
<i>cagA</i> (5' end)‡	cagA1	GAT AAC AGG CAA CGT TTT CAG GGA	394	60, 1
	cagA2	CCG AAC GGA TCA AAA ATT CAT GG		
<i>cagA</i> (3' end)*‡	Ys	ACC CTA GTC GGT AAT GGG TTA	595	62, 1.5
	A6	GGC TGT TAG TAG CGT AAT TGT C		
<i>cagC</i> *	FURF	GCG TAA GCA AAA ACA GTC GCC TGA	412	51, 1.5
	RURF1	AAG AAA AAT GGT TGG GAC AAG TAG		
<i>cagC</i> †	FURF	See above	241	52, 1
	cagC-R	CAA GAA TCA CTG ACA GCT ACA AG		
<i>virB4/cagF</i> *	CRB	CAC TCT CAA TGA ACC CGT TAT GAA	2,675	50, 1.5
	5B10	CTT TGT CAA GTC ATT TCT CAG		
<i>virB4</i> †	CRB	See above	594	60, 1
	cagE-R	CTG CCT AGC GTA ATA TCA CC		
<i>cagF</i> †	5B10	See above	261	60, 1
	cagF-R	CCT CGC TTA TGT TGT TAT TG		
<i>cagL/cagN/cagM</i> *	5BRB	TAT TGT CTG TTT TGA TGG CAG AAG	1,803	48, 1.5
	H12S	TGA TGA GCG ACA AAA CAA CTA TGC		
<i>cagL</i> †	5BRB	See above	354	60, 1
	cagL-R	CGG ATA TTC CGC ATT GTT GC		
<i>cagN</i> †	cagN-F	CGC CAC TAA CGC TTT GAG	277	60, 1
	cagN-R	CGG AGC AGC AGG TTT CAC		
<i>cagM</i> †	cagM-F	GCA TGG CGA TTG ATA AGA TC	254	60, 1
	cagM-R	CAT AGG ATC GCT ATC TAC G		
<i>cagP/cagQ/cagR</i> *	ARTF2	TAT TGC CTC GTT GAT CAA ACA AAG CTC TGC	1,327	47, 1.5
	HR1	GAT TTG GAG TTA GAT TAG GGT GGT		
<i>cagP</i> †	cagP-F	AGC CTT TAT TAT AGG CTG TTC	260	58, 1
	cagP-R	AAC CAA TTT TGC CAT TGA GTC		
<i>cagQ</i> †	cagQ-F	TGC TTC CTA CTA AAA CAC GC	309	58, 1
	cagQ-R	CAA CCA AAG CAG ATC CCA TG		
<i>cagS/virB7</i> *	B2L1	AAG TGC ATG CAG TAA TTA TGC GAA	1,627	47, 1.5
	E64Q	TCT AAA GAG AAA CCA AAT CCA TTG		
<i>cagS</i> †	cagS-F	GGG AGC GTT AGA TAA GGT TC	279	58, 1
	cagS-R	GTG GGA GCT TAG TGC CAT AC		
<i>virB7</i> †	cagT-F	CTG ATA GAG AAG TAT AGT GAG	337	58, 1
	cagT-R	GCA ACA CAT CAA AGG TCT G		
<i>virB9</i> *†	HP528F	GTT GGC TGT TTC TGT CTT GG	339	60, 1
	HP528R	GTA ATC TCT TGT CAT TAG GGC		
<i>virB10</i> *†	HP527F	GAT GAA ATC CAA ATA AGG CAA G	376	60, 1
	HP527R	CCA CTT GAA CTT TTT GTT GG		
<i>virD4/virB11</i> *	HP525F1	TAC CTA GCA AAA CAA CCT ATC CTC	1,955	48, 1.5
	HP525R1	TCC ACA TGT TTC TAG TAG CAG GAC		
<i>virB11</i> *†	HP525F2	ATG ACT GAA GAC AGA TTG AG	506	58, 1
	HP525R2	GCA ATA CCA TCT TTA ATC GC		
<i>virD4</i> *†	HP524F2	CTT TGG CTT TTG GTT GCA AG	513	58, 1
	HP524R2	CTA GTA GGA GCA ATC AAG C		
<i>IS605</i> *	H12H	TCA TCG CTC CAT AAC TAG C	1,486	54, 1.5
	E64U	GGT GGC TAT CTC ATC TAT AAC AGA		
Empty site‡	HP519	GCT TGC TTG TAT TGG CCT TG	324	58, 1
	HP549	GCA TGC ACA TTC CCT AAA GTG		
<i>JHP rrn16S</i> *†	HS1	GTG CTT ATT CGT TAG ATA CCG TCA T CTA GCA AGC	399	55, 1
	HS2	TAG AAG CTT CAT CGT T		

<sup>a</sup> \*, Primers used for PCR amplification (direct analysis of *cag* PAI in clinical strains by PCR and probe obtention for dot blot hybridization). The sequence of these primers was derived from the published sequence of *cag* PAI region of *H. pylori* NCTC 11638 (GenBank accession numbers U60176, U60177, and AC000108) (2, 5). †, Primers used for RT-PCR analysis of *cag* PAI genes; their sequences were derived from the published sequence of the *H. pylori* strains NCTC 11638 (GenBank accession numbers U60176, U60177, and AC000108) (2, 5), J99 (<http://scriabin.astrazeneca-boston.com/hpylori>), and 26695 (<http://www.tigr.org>). ‡, Primers used for PCR amplification and sequencing.

<sup>b</sup> Annealing step for each cycle. The cycling programs were preceded by one step at 94°C for 5 min; 35 cycles were applied; each consisted of 1 min at 94°C, annealing as specified in the table, and 72°C for a time dependent on the expected product size (1 min per kb). Finally, an elongation step at 72°C for 7 min was added.

confirmed by this technique. Dot blot analysis were then performed on all strains with all of the probes. An example of detection of the *cagL/cagM/cagN* region and *IS605* by hybridization is illustrated in Fig. 1B with the same strains as used for PCR detection (Fig. 1A). The same results were obtained for

all of these strains except for strain SS1, which gave positive results by dot blot and negative results by PCR. With the *IS605* probe, two contradictory results were obtained with DNA from strains in the lines 5 and 14. In summary, the correlation between the results obtained by PCR and those obtained by

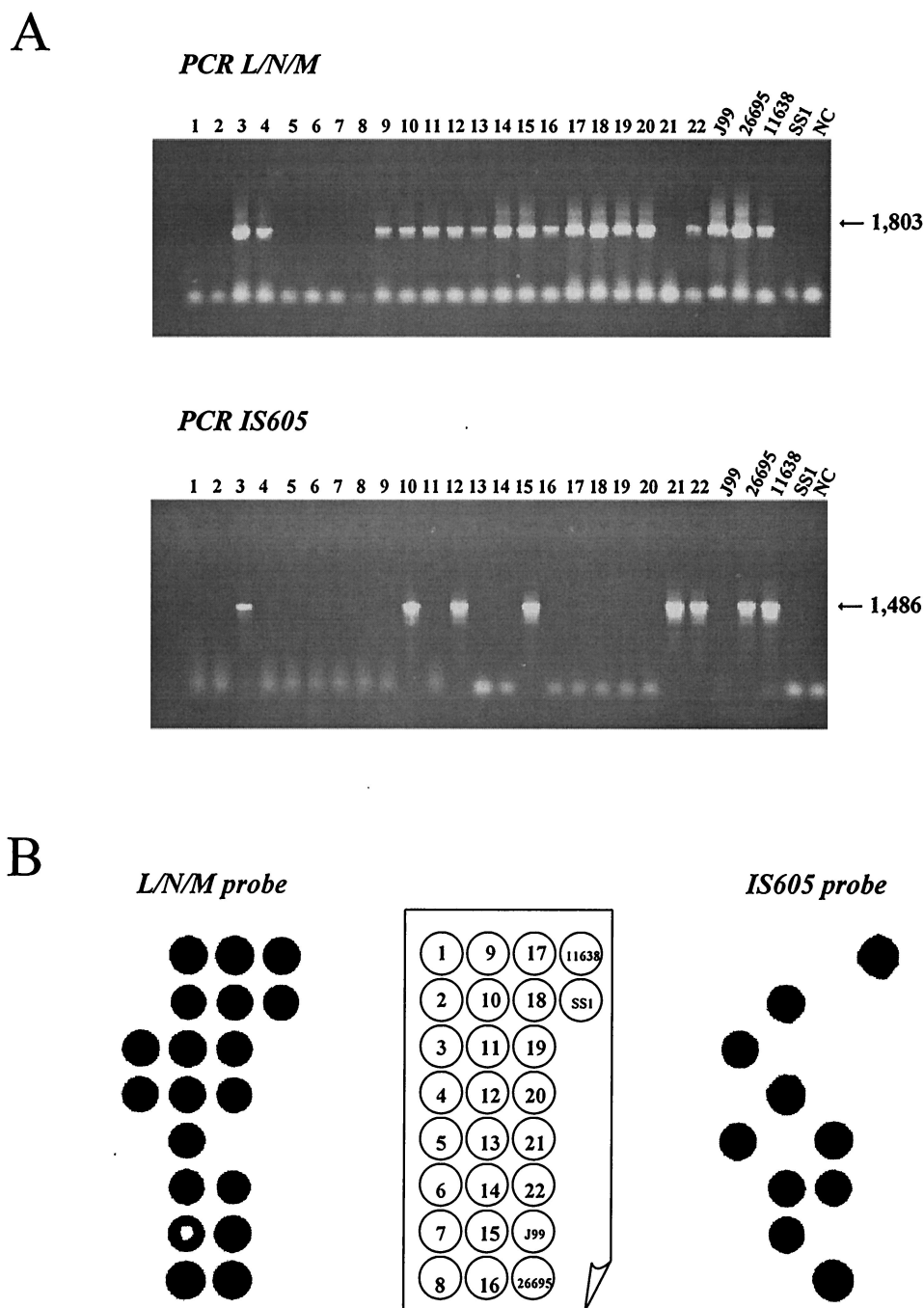


FIG. 1. Detection of the *cagL/cagN/cagM* region and *IS605* by PCR (A) and dot blot hybridization (B) for 22 clinical *H. pylori* strains and for the 4 reference strains. A schematic representation of DNA isolate positions is in the center. The sizes of amplified fragments are indicated in base pairs. NC, negative control.

dot blot were 100% for *cagC*, *virB4/cagF*, *cagS/virB7*, *virB9*, *virB10*, *virB11*, and *virD4*; 97.9% for *cagA* D008-R008, *cagA* Ys-A6, and *cagL/cagM/cagN*; and 95.7% for *cagP/cagR* and *IS605*. A strain was defined as lacking a given region if the results obtained both by PCR and by hybridization were negative. In the case of a given region detected by hybridization and not by PCR, the strain was considered positive for this region. The reverse case (positive PCR and negative hybridization) was never observed.

According to the rule established above, the results were analyzed first globally and then according to the associated disease. Comparison of open reading frame (ORF) patterns obtained from the different strains was performed using the unweighted-pair-group method with averages (30). The distribution of the selected *cag* PAI regions among strains is presented in Fig. 2. Each strain was assigned to a pattern based on the presence or absence of the selected region. The strains could be classified into four groups. The largest group included

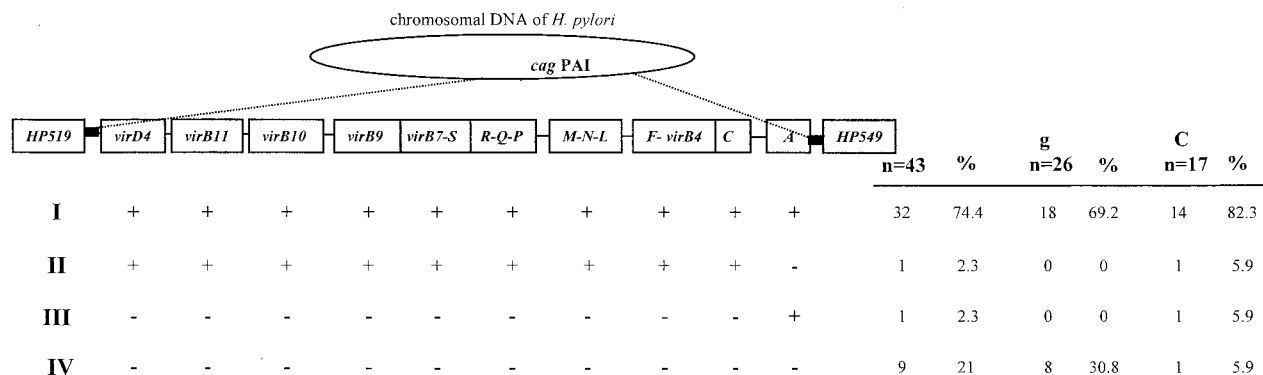


FIG. 2. Distribution of the *cag* PAI selected regions in the 43 *H. pylori* strains isolated from patients with chronic gastritis-associated dyspepsia (g) or gastric carcinoma (C). The four different patterns are presented in the left of the figure. The status for each selected region is presented (each pattern showed positive [+] or negative [-] hybridization with each probe). The different probes, as well as the two flanking genes, HP519 and HP549, are indicated at the top. The boundaries of the *cag* PAI are indicated by black boxes. The numbers and the percentages of clinical *H. pylori* strains presenting the different patterns are indicated at the right of the figure.

32 strains (74.4%) which were positive for all of the *cag* PAI regions studied. One strain (2.3%) exhibited a deletion of the *cagA* gene while the other regions were present, and one *cagA*-positive strain (2.3%) lacked all other regions, indicating that an association between *cagA* and *cag* PAI, while very strong, is not absolute. The fourth group included nine strains (21%) which contained none of the *cag* PAI regions studied. The SS1 reference strain was found to be positive for all of the regions studied (type I strain).

In order to confirm the complete absence of *cag* PAI in the nine strains, an amplification was performed using two primers (Table 1) located in the two genes flanking the *cag* PAI in the reference strains (HP519 and HP549; Fig. 2). As expected from the literature (1, 14), a single 324-bp amplicon was obtained from each of the nine *cag* PAI-negative strains. The nucleotide sequences of the so-called "empty site" were determined and deposited in GenBank under the accession numbers AF289390 to AF289398. The nine strains considered to be *cag* PAI negative based on PCR and dot blot results were, therefore, confirmed to be truly *cag* PAI negative.

The results are also presented according to the associated disease (Fig. 2). The proportion of each expression profile was compared between strains isolated from patients with gastric carcinoma and patients with chronic gastritis-associated dyspepsia using the chi-square test. Strains with no *cag* PAI regions were found more frequently among gastritis strains than among carcinoma strains, but the difference was not statistically significant (30.8 versus 5.9%, respectively,  $P = 0.11$ ). At the same time, strains containing all of the tested *cag* PAI regions were found more frequently among the gastric carcinoma strains (82.3%) than among the gastritis strains (69.2%). However, this difference was still not significant ( $P = 0.28$ ). Profiles II and III were not considered in this analysis because they did not contain enough strains ( $n = 1$  in each group).

Concerning the results obtained for IS605, 11 strains were found to be positive (25.6%): 8 in the gastritis group (30.8%) and 3 in the carcinoma group (17.6%). Although IS605 seemed to be more frequently found in strains isolated from patients with gastritis than in those isolated from patients with gastric cancer, this difference was not significant ( $P = 0.54$ ).

**Analysis of the *cagA* terminal sequences.** The sequence of two fragments of the *cagA* gene (at the 5' and 3' ends) was analyzed for the 33 *cagA*-positive strains. The 5' end amplicons were identical in size (394 bp), and their nucleotide sequences exhibited an overall similarity of 96.8% (accession numbers AF289399 to AF289413 for the gastric carcinoma strains and accession numbers AF289432 to AF289431 for the gastritis strains). In contrast, an interstrain size variation of the Ys-A6 fragment of the *cagA* gene (3' end) was observed (from 570 to 680 bp). Each amplicon was therefore submitted to sequencing (accession numbers AF289432 to AF289446 for the carcinoma strains and accession numbers AF289447 to AF289464 for the gastritis strains). The deduced amino acid sequences obtained were analyzed according to the method of Evans et al. (9). Eleven different sequence types were obtained, indicating a large variability. Regarding the number of repeated sequences, 23 strains contained only one repeat (69.7%) and 10 strains contained two repeats (30.3%). No example of three repeats was found. The number of repeats was analyzed according to the origin of the strains. Although two repeated sequences were found more frequently in strains from patients with gastric cancer (40 versus 22.2%), this difference was not significant ( $P = 0.46$ ).

**Analysis of TPMs of the CagA protein.** To determine whether the CagA protein of the clinical strains was phosphorylated, putative CagA TPMs were sought in their sequences. Three sites of phosphorylation have been recently identified by Odenbreit et al. (25): site "A" (KFGDQRY) at amino acid (aa) 122, site "B" (KNSTEPY) at aa 899, and site "C" (KLKD STKY) at aa 1029. Thus, the sequences of the two fragments of the *cagA* gene which may contain these three sites (TPM A in the A1-A2 fragment and TPMs B and C in the Ys-A6 fragment) were analyzed. The *cagA* sequences of all 33 Costa Rican strains have the TPM KFGDQRY in position A. Nineteen strains (12 of 15 patients with gastritis and 7 of 18 patients with gastric carcinoma) presented the TPM KNST/GEPIY in position B, the other strains having a variable sequence which did not correspond to a phosphorylation motif (KNEPIY, 12 strains; ENSAEPIY, two strains). No strain had a TPM in position C. Thus, globally, the majority of the strains (57.6%)



was similar to the 26695 strain which contains TPMs A and B. The remaining strains contained only the TPM A site, as did the strain NCTC 11638. In conclusion, the CagA protein from all of the Costa Rican strains tested contained at least one TPM; however, no association was found between the number of TPMs and the type of disease from which the strain originated. Indeed, 66.7% of the strains from patients with chronic gastritis only presented two TPMs (A and B) in the CagA protein versus 46.7% of the gastric carcinoma strains ( $P = 0.24$ ).

**Expression of the *cag* PAI genes in *H. pylori* strains.** To investigate whether the ORFs present in the *cag* PAI regions were expressed in vitro, transcription of 15 genes of *cag* PAI (*cagA*, *cagC*, *virB4*, *cagF*, *cagL*, *cagN*, *cagM*, *cagP*, *cagQ*, *cagS*, *virB7*, *virB9*, *virB10*, *virB11*, and *virD4*) was investigated by reverse transcription using a random primer mixture followed by a PCR (Enhanced Avian RT-PCR Kit; Sigma Aldrich, St. Quentin Fallavier, France) with specific primers (Table 1). RNA from an in vitro *H. pylori* culture was isolated according to the previously described method (24). Among the 15 genes, only one was found to be expressed in all of the strains under the conditions tested (*cagF*). The expression of five genes (*cagC*, *cagP*, *cagQ*, *virB11*, and *virD4*) was never detected. The other genes ( $n = 9$ ) were differentially transcribed among the *cag* PAI-positive strains: *cagL* in only 1 strain (3%), *cagS* and *virB4* in 4 strains (12.1%), *cagN* and *virB9* in 5 strains (15.1%), *virB7* in 7 strains (21.2%), *cagM* in 11 strains (33.3%), *virB10* in 21 strains (63.6%), and *cagA* in 29 strains (87.9%). No correlation could be drawn from this analysis. However, two pairs of genes seemed to be linked: *cagP* and *cagQ* on the one hand and *virB11* and *virD4* on the other hand. Indeed, when the *cagP* gene was not expressed, *cagQ* was not expressed either. The same observation was noted for the *virB11* and *virD4* genes. This fact, together with their relative locations (Fig. 2), suggested that these pairs of genes belong to the same transcriptional unit and may be coregulated. The expression status of genes was analyzed according to the disease of the patients from whom the strains were isolated (Table 2). The only *cag* PAI gene for which a statistically significant difference was found ( $P = 0.03$ ) was the *virB10* gene, whose transcript was detected in 21 strains (48.8%), among which 9 were isolated from gastritis patients (34.6%) and 12 were isolated from carcinoma patients (70.6%). A trend, while not statistically significant, was also observed for *cagN* ( $P = 0.07$ ) and *cagM* ( $P = 0.08$ ) expression.

The first question posed in this study was whether the development of gastric carcinoma in Costa Rica, known to display a high incidence of this pathology (29), is associated with the presence of *cag* PAI in infecting *H. pylori* strains. The results of our study showed that, in this country, most of the *H. pylori* strains are positive for the *cag* PAI (74.4%). Nevertheless, while no statistically significant difference was found, eight of the nine strains with no *cag* PAI were isolated from patients with gastritis but only one was isolated from a gastric carcinoma patient, indicating a tendency ( $P = 0.11$ ). There are controversial reports concerning the association of the *cag* PAI-positive *H. pylori* and more severe gastric diseases such as peptic ulcer disease and gastric carcinoma. Most of the reports from Western countries (Europe and the United States) (5, 13) found a positive association, while no association was found in

TABLE 2. Distribution of expressed genes of the *cag* PAI between gastritis and gastric carcinoma strains

Gene <sup>a</sup>	No. of genes ( $n = 43$ )	% Total	Gastritis patient isolates		Gastric carcinoma patient isolates		$P^b$
			No. ( $n = 26$ )	%	No. ( $n = 17$ )	%	
<i>cagA</i> *	29	67.4	16	61.5	13	76.5	0.3
<i>cagC</i> †	0	0	0	0	0	0	
<i>virB4</i> *	4	9.3	3	11.5	1	5.9	0.6
<i>cagF</i> *	34	79.1	18	69.2	15	94.1	0.1
<i>cagL</i> *	1	2.3	0	0	1	5.9	0.4
<i>cagN</i> *	5	11.6	1	3.8	4	23.5	0.07
<i>cagM</i> ‡	11	25.6	4	36.4	7	63.6	0.08
<i>cagP</i> †	0	0	0	0	0	0	
<i>cagQ</i> †	0	0	0	0	0	0	
<i>cagS</i> *	4	9.3	4	15.4	0	0	0.1
<i>virB7</i> *	7	16.3	4	15.4	3	17.6	1
<i>virB9</i> *	5	11.6	5	19.2	0	0	0.1
<i>virB10</i> *	21	48.8	9	34.6	12	70.6	0.03
<i>virB11</i> †	0	0	0	0	0	0	
<i>virD4</i> †	0	0	0	0	0	0	

<sup>a</sup> \*, Genes expressed in the J99 and 26695 strains; †, genes not expressed in the J99 and 26695 strains; ‡, gene expressed in the J99 strain but not in the 26695 strain.

<sup>b</sup> Data obtained by using Fisher's exact test.

the Far East (23). This fact can be explained by the high background level of *cag* PAI-positive *H. pylori* strains in far eastern countries which prohibits making conclusions based on cross-sectional or case-control studies. It is also possible that, while all *H. pylori* strains lead to cancer if the patient lives long enough, the process may develop more quickly in certain individuals due to environmental and individual genetic factors. However, concerning the absence of the *cag* PAI in strains isolated from patients with more severe disease (one case in the present study), it cannot be excluded that the strains studied, which were present at the time of carcinoma diagnosis, were not necessarily the carcinogenic strains, since gastric carcinoma evolves over several decades. Indeed, the possibility of harboring several strains successively, especially in a country with a high prevalence of infection, does exist (22, 23, 28). Despite the fact that *cag* PAI is present in most of the strains, it may be that the composition of the *cag* PAI genes or the regulation of their expression varies and influences the associated pathogenicity.

Therefore, in the present study, the composition of the *cag* PAI was analyzed in depth. The regions sought were chosen based on their location on the island as well as on the putative role of the ORFs encoded (2, 4, 5, 25, 27, 31, 34). Our results did not show an association between a particular region and the disease status: in 32 *H. pylori* clinical strains, 10 different regions of the *cag* PAI corresponding to 16 ORFs were all present, and therefore the strains could be categorized as type I according to Censini et al. (5). Interestingly, in one strain, the *cagA* gene itself was absent, whereas the other regions of the *cag* PAI were present, indicating that the use of this gene as the sole marker for the detection of the *cag* PAI may sometimes lead to erroneous results. Furthermore, in another strain, isolated from a gastric carcinoma patient, the *cagA* gene was the only part of the *cag* PAI present, leading us to the question of what are the pathogenic properties of this strain because the

type IV secretion system which delivers the CagA protein into the epithelial cells is encoded by genes which are absent.

The variation of the CagA protein size has been related to the presence of a variable number of repeated sequences in the 3' end of the gene (9). The presence of these multiple repeats in the COOH terminus of the protein was found to be associated with the development of gastric carcinoma in Japanese (36), U.S., and Colombian *H. pylori* strains, but not in Korean strains (35). In our study on Costa Rican strains, it was not possible to confirm this association.

In a recent study, Odenbreit et al. (25) studied the tyrosine phosphorylation of the CagA protein in several *H. pylori* reference strains and established that the translocation of the CagA protein, as well as the presence of at least one of three TPMs in the CagA protein, is necessary for CagA phosphorylation; therefore, this characteristic may be relevant in terms of disease association. Thus, such motifs were sought in the deduced CagA protein sequence in the strains included in our study. The results show that half of them contained only one phosphorylation site (at position 122), such as strain NCTC 11638, whose CagA protein is weakly phosphorylated (25). The remaining strains presented two phosphorylation sites (at positions 122 and 899), such as strain 26695, which exhibited a moderate phosphorylation of the CagA protein. This dichotomy was independent of the clinical origin of the strain. It is possible that the CagA protein of all the Costa Rican strains is phosphorylated. These data could be confirmed by detection of phosphorylated tyrosine and translocated CagA proteins in epithelial cell cultures infected by these strains.

For the first time, we reported the *in vitro* expression of 14 *cag* PAI genes in addition to the *cagA* gene for all the clinical strains. Our results showed that the *cagA* gene is expressed *in vitro* in 87.9% of the strains. In a previous study, Maeda et al. (16) reported also the differential expression of the *cagA* gene in various *H. pylori* strains. Among the ORFs which were never expressed, two gene pairs, *cagP-cagQ* and *virB11-virD4*, were located in close proximity and were similarly orientated, indicating their possible inclusion in two independent operons. The genes whose encoded proteins form the secretion system IV were expressed differentially among the strains. The *virB10* gene was expressed more frequently in carcinoma than in gastritis-only strains ( $P = 0.03$ ). However, two of these genes were never expressed (*virB11* and *virD4*). It is possible that these genes, which are involved in the translocation of the CagA protein into gastric cells, are only expressed following the contact of the bacterium with the epithelial cell, and therefore their expression could not be detected *in vitro*.

The IS605 element was found in 11 of 43 isolates (26.5%). This proportion is similar to that found by Kersulyte et al. (14) using hybridization or PCR on a set of strains originating from patients also living in Latin America, including Peru (13 of 69, 19%) and Guatemala (16 of 44, 36%).

Despite an in-depth exploration of the composition of the *cag* PAI and the known characteristics of what is now considered to be a major pathogenicity factor, i.e., the *cagA* gene, we were able to find only one element (the expression of the *virB10* homolog gene) which was associated with the severe outcome of *H. pylori* infection, i.e., gastric carcinoma. It seems that the *cag* PAI constitutes a group of genes encoding proteins which are highly complementary in their function since they

have been transmitted as a whole throughout evolutionary events subsequent to their incorporation.

Comparison of the genome of strains J99 and 26695 (3, 7, 17, 33) showed the presence of other regions whose role in pathogenicity remains to be explored. It is likely that virulence results from a combination of various factors, such as the presence of the *cag* PAI together with that of such regions as those encoding, cytotoxin production, and adhesin expression, rather than the presence of a unique virulence factor (6). Moreover, a study of the transcriptome under more physiological conditions and of the proteome may, in the future, provide information on additional elements for what remains the most important pathogenic determinant of *H. pylori*.

We acknowledge the financial support of the Association pour la Recherche sur le Cancer (ARC), Villejuif (no. 9313), France, and the Conseil Regional d'Aquitaine, Bordeaux, France. A. Occhialini is a fellowship recipient from IRMAD, Paris, France.

We gratefully acknowledge Fernando Garcia (San José, Costa Rica) for providing gastric biopsy samples. We also thank Kathryn Mayo (Laboratoire de Bactériologie-Université Bordeaux 2) for English corrections.

#### REFERENCES

- Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. J. Pan, S. Suerbaum, S. A. Thompson, A. Van der Ende, and L. J. van Doorn. 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* 32:459–470.
- Akopyants, N. S., S. W. Clifton, D. Kersulyte, J. E. Crabtree, B. E. Youree, C. A. Reece, N. O. Bukanov, E. S. Drazek, B. A. Roe, and D. E. Berg. 1998. Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol.* 28:37–53.
- Alm, R. A., L.-S. L. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176–180.
- Asahi, M., T. Azuma, S. Ito, Y. Ito, H. Suto, Y. Nagai, M. Tsubokawa, Y. Tohyama, S. Maeda, M. Omata, T. Suzuki, and C. Sasakawa. 2000. *Helicobacter pylori* CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *J. Exp. Med.* 191:593–602.
- Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA* 93:14648–14653.
- Covacci, A., S. Falkow, D. E. Berg, and R. Rappuoli. 1997. Did the inheritance of a pathogenicity island modify the virulence of *Helicobacter pylori*? *Trends Microbiol.* 5:205–208.
- Doig, P., B. L. de Jonge, R. A. Alm, E. D. Brown, M. Uria-Nickelsen, B. Noonan, S. D. Mills, P. Tummino, G. Carmel, B. C. Guild, D. T. Moir, G. F. Vovis, and T. J. Trust. 1999. *Helicobacter pylori* physiology predicted from genomic comparison of two strains. *Microbiol. Mol. Biol. Rev.* 63:675–707.
- El-Omar, E. M., M. Carrington, W. H. Chow, K. McColl, J. H. Bream, H. A. Young, J. Herrera, J. Lissowska, C. C. Yuan, N. Rothman, G. Lanyon, M. Martin, J. F. Fraumeni, Jr., and C. S. Rabkin. 2000. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* 404:398–402.
- Evans, D. J., D. M. Queiroz, E. N. Mendes, and D. G. Evans. 1998. Diversity in the variable region of *Helicobacter pylori* *cagA* gene involves more than simple repetition of a 102-nucleotide sequence. *Biochem. Biophys. Res. Commun.* 245:780–784.
- Huang, J. Q., S. Sridhar, Y. Chen, and R. H. Hunt. 1998. Meta-analysis of the relationship between *Helicobacter pylori* seropositivity and gastric cancer. *Gastroenterology* 114:1169–1179.
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. 1994. *Helicobacter pylori*, p. 177–240. In Schistosomes, liver flukes and *Helicobacter pylori*: views and expert opinions of an IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. IARC, Lyon, France.
- Ikeno T., H. Ota, A. Sugiyama, K. Ishida, T. Katsuyama, R. M. Genta, and S. Kawasaki. 1999. *Helicobacter pylori*-induced chronic active gastritis, intestinal metaplasia, and gastric ulcer in mongolian gerbils. *Am. J. Pathol.* 154:951–960.
- Jenks, P. J., F. Mégraud, and A. Labigne. 1998. Clinical outcome after infection with *Helicobacter pylori* does not appear to be reliably predicted by the presence of any of the genes of the *cag* pathogenicity island. *Gut* 43:752–758.

14. Kersulyte, D., H. Chalkauskas, and D. E. Berg. 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Mol. Microbiol.* **31**:31–43.
15. Lee, A., J. O'Rourke, M. C. De Ungria, B. Robertson, G. Daskalopoulos, and M. F. Dixon. 1997. A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney Strain. *Gastroenterology* **112**:1386–1397.
16. Maeda, S., H. Yoshida, T. Ikenoue, K. Ogura, K. Kanai, N. Kato, Y. Shiratori, and M. Omata. 1999. Structure of *cag* pathogenicity island in Japanese *Helicobacter pylori* isolates. *Gut* **44**:336–341.
17. Marais, A., G. L. Mendz, S. L. Hazell, and F. Mégraud. 1999. Metabolism and genetics of *Helicobacter pylori*: the genome era. *Microbiol. Mol. Biol. Rev.* **63**:642–674.
18. Marshall, B. J., and J. R. Warren. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **i**:1311–1315.
19. McFarlane, G. A., and A. Munro. 1997. *Helicobacter pylori* and gastric cancer. *Br. J. Surgery* **84**:1190–1199.
20. Mégraud, F. 1996. Advantages and disadvantages of current diagnostic tests for the detection of *Helicobacter pylori*. *Scand. J. Gastroenterol.* **31**:57–62.
21. Mégraud, F., and H. Lamouliatte. 1992. *Helicobacter pylori* and duodenal ulcer. Evidence suggesting causation. *Dig. Dis. Sci.* **37**:769–772.
22. Morera-Brenes, B., R. Sierra, R. Barrantes, J. Jonasson, and C. E. Nord. 1994. *Helicobacter pylori* in a Costa Rican dyspeptic patient population. *Eur. J. Microbiol. Infect. Dis.* **13**:253–257.
23. Mukhopadhyay, A. K., D. Kersulyte, J. Y. Jeong, S. Datta, Y. Ito, A. Chowdhury, S. Chowdhury, A. Santra, S. K. Bhattacharya, T. Azuma, G. B. Nair, and D. E. Berg. 2000. Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *J. Bacteriol.* **182**:3219–3227.
24. Occhialini A., A. Marais, R. Alm, F. Garcia, R. Sierra, and F. Mégraud. 2000. Distribution of open reading frames of plasticity region of strain J99 in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infect. Immun.* **68**:6240–6249.
25. Odenbreit, S., J. Püls, B. Sedlmaier, E. Gerland, W. Fischer, and R. Haas. 2000. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* **287**:1497–1500.
26. Queiroz, D. M., E. N. Mendes, G. A. Rocha, A. M. Oliveira, C. A. Oliveira, P. P. Magalhaes, S. B. Moura, M. M. Cabral, and A. M. Nogueira. 1998. *cagA*-positive *Helicobacter pylori* and risk for developing gastric carcinoma in Brazil. *Int. J. Cancer* **78**:135–139.
27. Segal, E. D., J. Cha, J. Lo, S. Falkow, and L. S. Tompkins. 1999. Altered states: Involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* **96**:14559–14564.
28. Sierra, R., N. Muñoz, A. S. Peña, I. Biemond, W. van Duijn, C. B. Lamers, S. Teuchmann, S. Hernandez, and P. Correa. 1992. Antibodies to *Helicobacter pylori* and pepsinogen levels in children from Costa Rica: comparison of two areas with different risks for stomach cancer. *Cancer Epidemiol. Biomark. Prev.* **1**:449–459.
29. Sierra, R., D. M. Parkin, and G. M. Leiva. 1989. Cancer in Costa Rica. *Cancer Res.* **49**:717–724.
30. Sneath, P. H. A., and R. R. Sokat. 1973. Numerical taxonomy: the principles and practice of numerical classification. W. H. Freeman & Co., San Francisco, Calif.
31. Stein, M., R. Rappuoli, and A. Covacci. 2000. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after *cag*-driven host cell translocation. *Proc. Natl. Acad. Sci. USA* **97**:1263–1268.
32. Sugiyama, A., F. Maruta, T. Ikeno, K. Ishida, S. Kawasaki, T. Katsuyama, N. Shimizu, and M. Tatematsu. 1998. *Helicobacter pylori* infection enhances *N*-methyl-*N*-nitrosourea-induced stomach carcinogenesis in the mongolian gerbils. *Cancer Res.* **58**:2067–2069.
33. Tomb, J.-F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539–547.
34. Walsh, J. H. 2000. CagA protein from *Helicobacter pylori* is a trojan horse to epithelial cells. *Gastroenterology* **118**:817–818.
35. Yamaoka, Y., H. M. T. El-Zimaity, O. Gutierrez, N. Figura, J. K. Kim, T. Kodama, K. Kashima, and D. Y. Graham. 1999. Relationship between the *cagA* 3' repeat regions of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology* **117**:342–349.
36. Yamaoka, Y., T. Kodama, K. Kashima, D. Y. Graham, and A. R. Sepulveda. 1998. Variants of the 3' region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. *J. Clin. Microbiol.* **36**:2258–2263.

Editor: A. D. O'Brien